Analyzing the Impact of Move-In Week on the Greater Boston Area

Brooke Mullen, Claire Russack, Dharmesh Tarapore, Vincent Wahl
bemullen@bu.edu, crussack@bu.edu, dharmesh@bu.edu, vinwah@bu.edu

***Abstract*—Move in week is universally chaotic and a generally unpleasant experience for both, college students and local residents [1]. Despite universities' meticulous attempts to mitigate trouble, the logistical concerns presented by the sudden influx of over 30,000 students are inevitable. In this project, we attempt to identify specific factors affecting residents' quality of life (as a result of move-in week) and study ways to improve it.**

I. Introduction

Over 50,000 new and returning students move in to dormitories and apartments in the city of Boston each year [2]. Known colloquially as "move-in week", this 7-day period between late August and early September wreaks considerable havoc over residents' lives. Traffic delays, overcrowded public transport, inadequate parking spaces, and the infamous *"Storrowed"* trucks [3] are among some of the difficulties that irk both, students and locals.

Due to the nature of an academic calendar for most universities, the lease cycle tends to be September 1st. This rapid movement in a dense area affects the overall quality of life in the city, often for the worse. [TODO: explain why we care(answering 2 questions...)].

Mensurability notwithstanding, conventional proxies[1] for quality of life disregard (among other things) societal happiness: a critical adjunct to traditional economic metrics [4]. To remedy this, we propose an approach that utilizes open data sourced from Boston City's *Analyze Boston* data portal [5] and posts made on the microblogging platform, Twitter [6]. In particular, we use the *CityScore*, *311 Service Requests*, and *Boston Fire Incident Reporting* datasets to identify the demand for specific city resources across multiple time periods as a function of user satisfaction.

The rest of this paper is organized as follows: in section II, we document related research that inspired this project. In section III, we describe each of the 4 chosen datasets in greater detail to focus on their merits, potential limitations, and the methods we used to discern their relative importance in estimating the city's quality of life. Section IV outlines our results and examines their implications. Section V concludes the report with an outlook on future research.

II. Related Work

Jim Haddadin explored the relationship between move-in week and garbage disposal concerns in [7]. Noting a sharp rise in the number of code violations around university residences during move-in week, Haddadin highlights some unforeseen consequences of improper trash disposal. In [4], Dodds et al address the subjectivity and vagueness inherent to estimating happiness among people by mining over 46 billion tweets to uncover temporal variations in happiness and information levels over timescales ranging from hours to years.Their remote-sensing 'hedonometer' algorithm generated a rich source of information about short-term, experiential happiness in a population and its causes. We use a similar, albeit slightly modified approach to understand residents' moods during move-in week.

III. Approach

Our desire to allay the worst aspects of move-in week presupposes the existence of a correlation between students moving in and a perceptible change in the quality of residents' lives. This requires:

i. The identification of metrics that can effectively approximate quality of life and

ii. Predictive analyses that capture and justify the impact of altering select attributes of move-in week[2].

Similar to the approaches charted in [4] and [7], we begin first by selecting datasets that might evince the impact of students moving in. Boston City's Open Data Initiative, *Analyze Boston* portal provides a vast repository of highly granular information about the city's functioning. In particular, the *311 Service Requests*, *CityScores*, and *Fire Incident Reporting* datasets are of particular interest and we expound their significance below [5]. For quality of life, we rely on the seminal research conducted by Mitchell et al in [8] and use geotagged Twitter posts within 50 kilometers of Boston city to gauge happiness. Limitations stemming from a lack of representativeness and potential bias in dataset selection are explored further in section IV.

A. Vincent

B. Claire

C. Library Attendance Analysis

The goal of analyzing library visits to city score started out with questioning whether the number of visits actually went up during university session. The two datasets, were filtered based

[1]The GDP (Gross Domestic Product) is often used as an approximate measure of quality of life, in conjunction with other metrics such as the HDI (Human Development Index) and Gini Coefficient.

[2]Note that this does **not** insinuate the existence of a causal relationship between university students moving in and residents' quality of life: it merely seeks to leverage potential correlations between the two.

on the "ETL_LOAD_DATE" feature from the City Score data set:

Students In Session

- September
- October
- November
- December 1st-15th
- January 16th-31st
- February
- March
- April
- May

Students Not In Session:

- January 1st-15th
- July
- June
- August
- December 16th-31st

D. Analyzing Quality of Life

Traditional economic indicators such as gross domestic product (GDP), Gini coefficient, and human development index (HDI) provide remarkable approximations for quality of life and are often indicative of a region's well-being and health [9][11]. In our analysis of Boston's quality of life, however, we avoid using these metrics for several reasons. First, the length of our timeframe of interest (7 days) renders the evolution of most economic indicators inconsequential. Second, given the geographic area of consideration (Greater Boston), none of the aforementioned metrics are appropriate[3]. Perhaps most importantly, economic prosperity does not provide an accurate representation of emotional well-being [11].

In contrast, Twitter provides a rich source of data which, when analyzed, can reveal societal-scale levels of happiness [8]. Twitter's erstwhile 140-character limit yields an extemporaneous dataset of user experiences [12], especially given most Twitter users' proclivity for posting in the moment [13][15].

Dodds et al are careful to draw a distinction between experential/fleeting happiness and contentment and stress that their 'hedonometer' algorithm efficiently estimates the former, while inferring the latter with increasing accuracy over large periods of time [4].

Comparing multiple text corpora using a single metric necessitates justifying the variation. A comprehensive explanation of the 'word shift' graphs is provided in [4]. We reproduce some of the calculations here:

Consider two texts T_{ref} (for reference) and T_{comp} (for comparison) with happiness scores $h_{avg}^{(ref)}$ and $h_{avg}^{(comp)}$, where the happiness score for a text is defined by:

$$h_{avg}(T) = \frac{\sum_{i=1}^{N} h_{avg}(w_i) f_i}{\sum_{i=1}^{N} f_i} = \sum_{i=1}^{N} h_{avg}(w_i) p_i \qquad (1)$$

where f_i is the frequency of the i^{th} word w_i for which we have an estimate of average happiness, $h_{avg}(w_i)$, and $p_i = \frac{f_i}{\sum_{j=1}^{N} f_j}$ represents the corresponding normalized frequency[4].

Using Eq (1), we can compare two texts thus:

$$h_{avg}^{(comp)-h_{avg}^{(ref)}} = \sum_{i=1}^{N} h_{avg}(w_i)[p_i^{(comp)-p_i^{(ref)}}]$$

$$= \sum_{i=1}^{N} [h_{avg}(w_i) - h_{avg}^{(ref)}][p_i^{(comp)-p_i^{(ref)}}] \qquad (2)$$

$$\because \sum_{i=1}^{N} h_{avg}^{(ref)}[p_i^{(comp)} - p_i^{(ref)}] = h_{avg}^{(ref)}[p_i^{(comp)} - p_i^{(ref)}]$$

$$= h_{avg}^{(ref)}(1-1) = 0$$

Two aspects determine the change in the i^{th} word's contribution:

1. Whether the word is (on average) happier than the reference text's average, $h_{avg}^{(ref)}$ and
2. The relative abundance of the word in T_{comp} to that in T_{ref}.

A word's happiness relative to T_{ref} is specified by $+$ (happier) and $-$ (less happier).

Relative abundance is specified similarly, using $\uparrow$ (more abundant) and $\downarrow$ (less abundant).

The combination of both results in the following four possibilities:

1. $+\uparrow$: Increased usage of relatively positive words.
2. $-\downarrow$: Decreased usage of relatively negative words.
3. $+\downarrow$: Decreased usage of relatively positive words.
4. $-\uparrow$: Increased usage of relatively negative words.

Eq (2) is normalized to obtain a normalized summand:

$$\delta h_{avg,i} = \frac{100}{|h_{avg}^{(comp)} - h_{avg}^{(ref)}|} t$$

where

$$t = \underbrace{[h_{avg}(w_i) - h_{avg}^{(ref)}]}_{+/-} \underbrace{[p_i^{(comp)} - p_i^{(ref)}]}_{\uparrow/\downarrow} \qquad (3)$$

where $\sum_i \delta h_{avg,i} = \pm 100$

A demonstration of this is provided in figure 1, using the *dharmSentiment* module [14]. Clearly, the negative words pressure, dead, broken, and trash were used more frequently, while the positive words joy, party, heart, and spring were used less frequently.

IV. Results

Results go here.

[3]GDP, HDI, and Gini coefficients are computed for countries as a whole and are thus unlikely to register meaningful changes sparked by move-in week in one city.

[4]The average happiness for over 10,000 words was calculated and made publicly available as part of the labMT dataset by Dodds et al

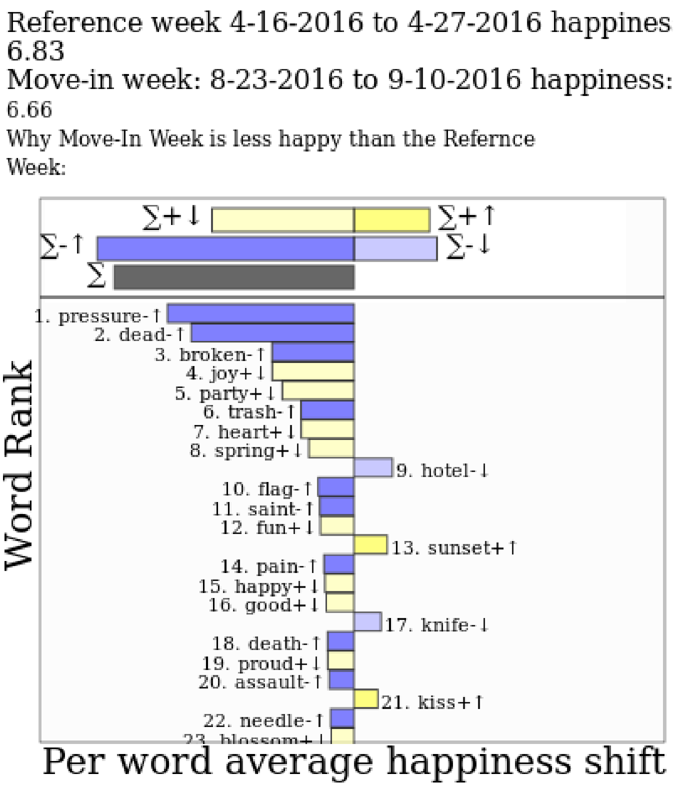

Fig. 1. A comparison of the happiness between move-in week and a randomly chosen week during the summer.

V. Conclusion

Our perfunctory investigation of the relationship between move-in week and quality of life is indicative of a negative correlation between move-in week and quality of life. While happiness (which was a heavily weighted component in our estimations of quality of life) is admittedly subjective and self-reported, we are confident that our approach to its evaluation provides a tenable, if somewhat inscrutable picture of happiness for the population as a whole. Time and primarily computational constraints prevented us from developing an inferential model of the shift in public happiness, given some of the optimizations suggested in [TODO: cite vincent's optimizations here], but we are hopeful that future work will explore this.
The work done in this paper would not have been possible without the 'hedonometer algorithm', or the *simpleLabMT* software package [10], both of which were designed in part by Dr. Andrew Reagan at the University of Vermont.

VI. Outlook

References

[1] Landry, Lauren. "Boston, You've Been Warned: The College Students Are Here.? Americaninno.com, 25 Aug. 2014, www.americaninno.com/boston/move-in-dates-for-boston-college-students-back-to-school-in-boston/.

[2] United States, Congress, Meade, Peter. "Boston by the Numbers: Colleges and Universities" Boston Redevelopment Authority. www.bostonplans.org/getattachment/1770c181-7878-47ab-892f-84baca828bf3 2011.

[3] Slane, Kevin. "Trillium Has a New Beer Named for Trucks That Get 'Storrowed'." Boston.com, The Boston Globe, 2 Feb. 2018, https://www.boston.com/culture/lifestyle/2018/02/02/trillium-has-a-new-beer-named-for-trucks-that-get-storrowed.

[4] Dodds, Peter Sheridan, et al. "Temporal patterns of happiness and information in a global social network: Hedonometrics and Twitter." PloS one 6.12 (2011): e26752.

[5] Analyze Boston, https://data.boston.gov/.

[6] Twitter, https://twitter.com.

[7] J. Haddidin, *Trash City: How Does Moving Week Impact the Quality of Life in Boston?*,Analyze Boston, *https://data.boston.gov/showcase/trash-city*

[8] Mitchell, Lewis, et al. "The geography of happiness: Connecting twitter sentiment and expression, demographics, and objective characteristics of place." PloS one 8.5 (2013): e64417.

[9] Diener, Ed. "Subjective well-being: The science of happiness and a proposal for a national index." American psychologist 55.1 (2000): 34.

[10] andyreagan/labMT-simple "Andyreagan/Labmt-Simple." GitHub. N. p., 2018. Web. 30 Apr. 2018.

[11] Kahneman, Daniel, and Angus Deaton. "High income improves evaluation of life but not emotional well-being." Proceedings of the national academy of sciences 107.38 (2010): 16489-16493.

[12] Edgeworth, Francis Ysidro. Mathematical psychics: An essay on the application of mathematics to the moral sciences. Vol. 10. Kegan Paul, 1881.

[13] Dunlap, Joanna C., and Patrick R. Lowenthal. "Tweeting the night away: Using Twitter to enhance social presence." Journal of Information Systems Education 20.2 (2009): 129.

[14] Tarapore, Dharmesh. "Weirdindiankid/Dharmsentiment." GitHub. 2018. Web. 30 Apr. 2018, URL https://github.com/weirdindiankid/dharmSentiment

[15] Bruns, Axel, and Jean E. Burgess. "The use of Twitter hashtags in the formation of ad hoc publics." Proceedings of the 6th European Consortium for Political Research (ECPR) General Conference 2011. 2011.

[16] Inside Twitter: An In-Depth Look Inside the Twitter World, Sysmos Resource Library. Available at http:// www.sysomos.com/insidetwitter/. Accessed April 12, 2018.

[17] S. Fox, K. Zickuhr, and A. Smith, Tech. Rep., Pew Internet & American Life Project (2006), accessed April 13, 2018, URL http://www.pewinternet.org/Reports/ 2009/17-Twitter-and-Status-Updating-Fall-2009. aspx.

[18] D. Kahneman and J. Riis, in The science of well-being, edited by F. A. Huppert, N. Baylis, and B. Keverne (Oxford University Press, Oxford, UK, 2005), pp. 285? 304

[19] M. A. Killingsworth and D. T. Gilbert, Science Magazine 330, 932 (2010)